Molecular adaptations of NADP-malic enzyme for its function in C_4 photosynthesis in grasses

Clarisa E. Alvarez[1,10], Anastasiia Bovdilova[2,3,10], Astrid Höppner[4,10], Christian-Claus Wolff[2,3],

Mariana Saigo[1], Felipe Trajtenberg[5], Tao Zhang[6,7], Alejandro Buschiazzo[5,8], Luitgard Nagel-Steger[6,7],

Maria F. Drincovich [1], Martin J. Lercher [3,9] and Veronica G. Maurino [2,3]*

[1]Centro de Estudios Fotosinteticos y Bioquimicos (CEFOBI-CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, University of Rosario, Rosario, Argentina.
[2]Plant Molecular Physiology and Biotechnology Group, Institute of Developmental and Molecular Biology of Plants, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.
[3]Cluster of Excellence on Plant Sciences, Düsseldorf, Germany.
[4]Center for Structural Studies, Hreinrich Heine University Düsseldorf, Düsseldorf, Germany.
[5]Laboratory of Molecular and Structural Microbiology, Institut Pasteur de Montevideo, Montevideo, Uruguay.
[6]Institut für Physikalische Biologie, Heinrich Heine University, Düsseldorf, Germany.
[7]Institut of Complex Systems, Structural Biochemistry (ICS-6), Jülich, Germany.
[8]Integrative Microbiology of Zoonotic Agents, Department of Microbiology, Institut Pasteur, Paris, France.
[9]Institute for Computer Science and Department of Biology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.
[10]These authors contributed equally: Clarisa E. Alvarez, Anastasiia Bovdilova, Astrid Höppner.
*e-mail: veronica.maurino@uni-duesseldorf.de

Abstract

In C_4 grasses of agronomical interest, malate shuttled into the bundle sheath cells is decarboxylated mainly by nicotinamide adenine dinucleotide phosphate (NADP)-malic enzyme (C_4-NADP-ME). The activity of C_4-NADP-ME was optimized by natural selection to efficiently deliver CO_2 to Rubisco. During its evolution from a plastidic non-photosynthetic NADP-ME, C_4-NADP-ME acquired increased catalytic efficiency, tetrameric structure and pH-dependent inhibition by its substrate malate. Here, we identified specific amino acids important for these C_4 adaptions based on strict differential conservation of amino acids, combined with solving the crystal structures of maize and sorghum C_4-NADP-ME. Site-directed mutagenesis and structural analyses show that Q503, L544 and E339 are involved in catalytic efficiency; E339 confers pH-dependent regulation by malate, F140 is critical for the stabilization of the oligomeric structure and the N-terminal region is involved in tetramerization. Together, the identified molecular adaptations form the basis for the efficient catalysis and regulation of one of the central biochemical steps in C_4 metabolism.

Main

The mechanism of C_4 photosynthesis is a complex trait that confers strongly reduced photorespiration and greater photosynthetic efficiency, accompanied by increased nitrogen and water use efficiency, in plants growing in temperate or hot climates and under high irradiation[1,2,3]. Despite its biochemical and anatomical complexity, C_4 photosynthesis emerged independently more than 60 times; it is particularly prevalent in grasses, where it evolved in approximately 18 lineages[4]. The evolution of the C_4 trait involved the functional specialization of bundle sheath cells (BSC) and the cellular relocation and molecular adaptation of photosynthetic and photorespiratory enzymes[3]. The adaptation of cellular activities to the C_4 syndrome regularly involved gene duplications followed by neofunctionalizations, as well as changes in the expression and biochemical properties of existing gene products; however, many details of these crucial evolutionary changes remain obscure[5].

The largest C_4 biochemical subtype, which includes the major monocot crops *Zea mays* (maize) and *Sorghum bicolor* (sorghum), uses a C_4-specific isoform of nicotinamide adenine dinucleotide phosphate (NADP)-dependent malic enzyme (C_4-NADP-ME) to decarboxylate malate in BSC chloroplasts[6]. Optimized kinetics and tight regulation of this enzyme are essential to C_4 photosynthesis, because coordination of C_4-NADP-ME decarboxylation and Rubisco carboxylation rates is imperative in maintaining an effective flux via the C_4 pathway. C_4-NADP-ME is likely to have evolved through the duplication of a gene encoding a plastidic non-photosynthetic isoform (nonC_4-NADP-ME), followed by the acquisition of a BSC-specific expression pattern[7,8,9,10]. In monocots, this gene duplication event occurred before the divergence of the Poaceae[5]. This means that the same duplicated NADP-ME was probably co-opted for the C_4 function in independent C_4 lineages[5].

Plastidic C_4- and nonC_4-NADP-MEs differ in their kinetic and regulatory properties. In maize and sorghum, C_4-NADP-ME shows two- to fivefold higher catalytic efficiency than that of the nonC_4-isoforms[11]. The affinity of C_4-NADP-ME for malate is twofold higher than that of the nonC_4-isoform[8,11]. In both species, the C_4-NADP-ME isoform is inhibited by malate at pH 7.0 (refs. [8,11,12,13,14]), and the affinity of C_4-NADP-ME for the co-factor NADP is two- to tenfold higher than that of the nonC_4-isoforms[8,14]. C_4- and nonC_4-NADP-ME also differ in their quaternary organization: the photosynthetic isoform assembles as a tetramer while the nonC_4 isoform assembles as a dimer[7,8,11,15]. The particular kinetic properties and oligomeric organization of C_4-NADP-ME seem to be crucial features for C_4 optimization, as they were also described for the independently evolved C_4-NADP-ME of *Saccharum officinarum*[13].

Why are the regulation of C_4-NADP-ME activity by malate and the formation of a tetramer C_4 considered adaptations? During the night, when photosynthesis is not active and stromal pH decreases to around 7.0–7.5 (ref. [16]), C_4-NADP-ME can be inhibited by its own substrate. This prevents an extreme decrease in malate content that ensures a rapid kick-start for Rubisco action at the beginning of the next day and avoids carbon starvation during the night[17,18]. In addition to metabolite regulation, it has been proposed that the tetrameric organization of C_4-NADP-ME forms the basis for a second level of regulation—a proportion of C_4-NADP-ME probably loses its quaternary structure during the night and adopts a less active or inactive lower oligomerization state[13]. However, experiments to confirm these regulatory mechanisms in planta are extremely difficult, in part because the molecular determinants of C_4-NADP-ME regulation are still not identified.

The structural and biochemical changes that occurred during the adaptation of C_4 enzymes to their central role in C_4 photosynthesis must have been acquired through specific amino acid substitutions. This was addressed in the case of *Flaveria* C_4 phosphoenolpyruvate carboxylase over a series of publications showing that only two single mutations are sufficient to switch the function from C_3 to C_4 (refs. [19,20,21,22,23]); a single mutation in the active site, A774S, is sufficient to confer low affinity on the substrate PEP, and a single mutation at the inhibitory site, R884G, is sufficient to reduce malate/aspartate inhibition. Here, we developed an algorithm to detect amino acids that are strictly differentially conserved in plastidic photosynthetic and non-photosynthetic isoforms within the clade Panicoideae of the family Poaceae. From these, we selected amino acids of probable functional importance based on the crystal structures of maize and sorghum C_4-NADP-ME, which represent the first crystal structures of malic enzyme reported for plants, complementing the collection of previously resolved structures of malic enzyme that use NADP from other sources such as human, pigeon and bacteria[24,25,26,27,28,29,30] (Supplementary Table 1). The combination of sequence analysis, structural analysis and protein engineering allowed us to identify molecular determinants that confer C_4 properties on NADP-ME, such as its tetrameric organization, high affinity to the substrate and pH-dependent inhibition by malate.

Results

Determination of the crystal structure of maize and sorghum C_4-NADP-ME

To aid in identifying amino acid residues that contribute to C_4 optimization of NADP-ME, we solved the crystal structures of C_4-NADP-ME of maize (ZmC_4-NADP-ME) at 2.2 Å resolution and of sorghum (SbC_4-NADP-ME) at 2.0 Å (see Table 1 for data collection and refinement statistics).

	ZmC_4-NADP-ME	**SbC_4-NADP-ME**
Data collection [a]		
Space group	*I* 21 21 21	*C* 1 2 1
Cell dimensions		
a, b, c (Å)	135.96 147.16 261.62	208.10 64.26 202.72
α, β, γ (°)	90 90 90	90 94 90
Resolution (Å)	48.83–2.20 (2.28–2.20)[b]	29.28–2.00 (2.03–2.00)[b]
R_{merge}	0.063 (0.472)	0.085 (0.585)
$I / \sigma I$	18.7 (3.51)	9.1 (2.1)
Completeness (%)	100 (99.0)	100 (100)
Redundancy	5.9 (6.0)	3.7 (3.6)
Refinement		
Resolution (Å)	48.83–2.20 (2.28–2.20)	29.28–2.00 (2.05–2.00)
No. reflections	132,069 (12,943)	181,074 (13,297)
$R_{\text{work}} / R_{\text{free}}$	0.166 (0.209) / 0.207 (0.264)	0.180 (0.211) / 0.210 (0.252)
No. atoms		
Protein	17,500	15,643
Ligand/ion	17	297
Water	729	1715
B factors		
Protein	40.13	32.42
Ligand/ion	46.00	41.46
Water	39.83	40.90
R.m.s. deviations		
Bond lengths (Å)	0.018	0.010
Bond angles (°)	1.83	0.99

[a] Data for ZmC_4-NADP-ME and SbC_4-NADP-ME were collected from one single crystal each.

[b] Values in parentheses relate to the highest-resolution shell.

The quaternary structure of ZmC4-NADP-ME displays a homo-tetrameric assembly composed of two dimers (monomers A and B and monomers C and D, respectively; Fig. 1a). For an in-depth description of the overall structure refer to Supplementary information and Supplementary Figs. 1 and 2. The dimers are tilted relative to each other at an angle of ~55° (Fig. 1b and Supplementary Video 1). Residues forming the tetramer interface are mainly located in domains A and C (Supplementary Table 2). SbC4-NADP-ME is likewise assembled as a tetramer (Fig. 1c), with an overall structure highly similar to that of ZmC4-NADP-ME (Supplementary information). The sorghum enzyme displays monomers A and B with unambiguous electron density, whereas monomers C and D display signs of substantial flexibility with weak density. Hereafter, we use monomers A and B for further structural comparisons. SbC4-NADP-ME shows an NADP molecule bound in the active sites of monomers A and B. In monomer B, we found additional electron density close to the nicotine amide moiety of NADP, which most likely represents a pyruvate as it is in an orientation and position similar to the pyruvate binding site previously reported for human mitochondrial NAD(P)-ME (PDB-1PJ3)28 (Fig. 1, Supplementary Video 1 and Supplementary Fig. 3).

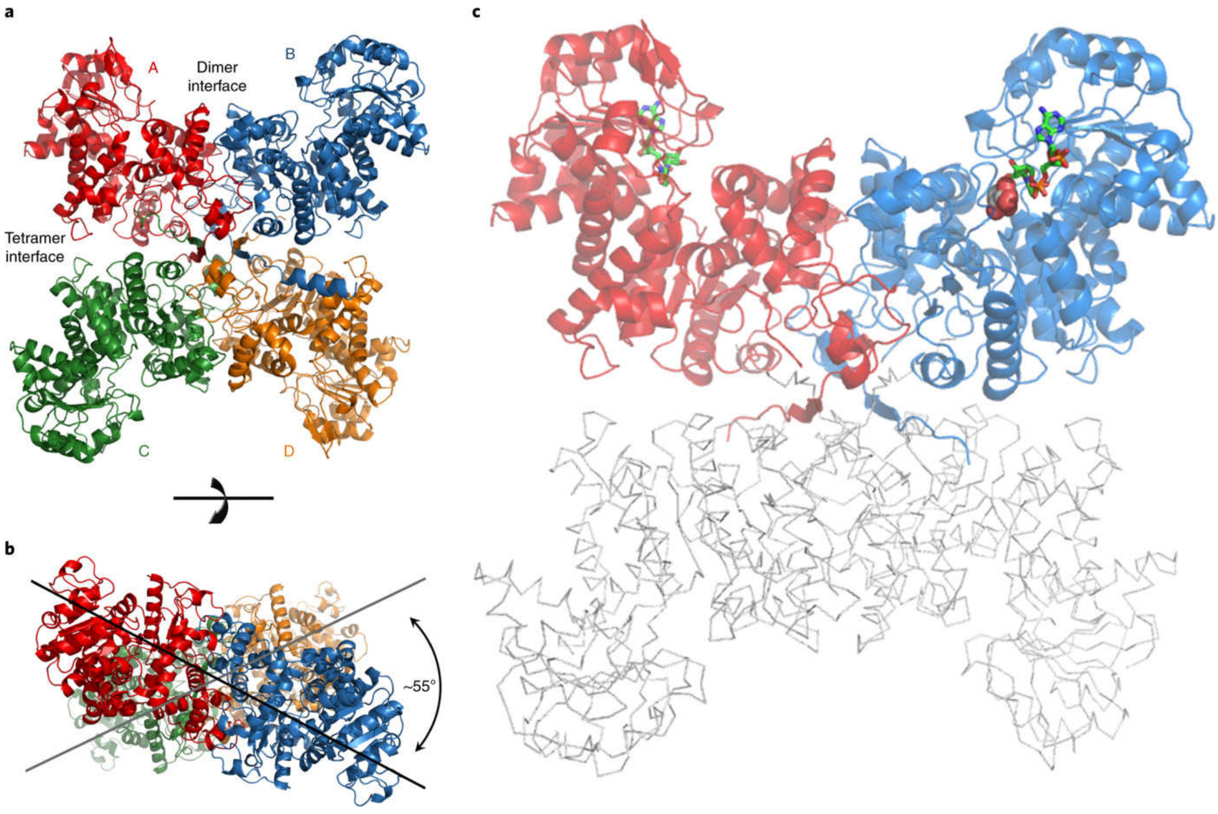


Fig. 1: Quaternary organization of C_4-NADP-ME a, Cartoon illustration of the tetrameric assembly of ZmC_4-NADP-ME. Monomer A (red) and monomer B (blue) represent the dimer; likewise, monomer C (green) and monomer D (orange). **b**, Top view of ZmC_4-NADP-ME, displaying the tilted arrangement of the dimers relative to each other. **c**, Cartoon representation of SbC_4-NADP-ME. Monomer A (red) with bound NADP (coloured sticks), monomer B (blue) with bound co-factor NADP (coloured sticks) and pyruvate (coloured spheres). Because large areas of monomers C and D are not well resolved by electron density, we depict these monomers as grey ribbons only to clarify the tetrameric

arrangement. Refer to Supplementary Video 1 for a three-dimensional (3D) overview of the overall structure of SbC_4-NADP-ME**.**

A superposition of ZmC4-NADP-ME monomers A and B indicates that the secondary structure elements from domains A and B adopt mainly the same relative spatial position, whereas domain C in monomer B is shifted outwards (Supplementary Fig. 4a; for details see Supplementary information). We therefore conclude that monomers A and D are in a 'closed' conformation while monomers B and C are in an 'open' conformation. In SbC4-NADP-ME, we found a very similar arrangement: monomer A adopts an 'open' conformation and monomer B a 'closed' conformation (root mean squared deviation (RMSD) monomers A and B: 0.988 Å over 509 C α atoms) (Supplementary Fig. 4b). The presence of these two conformations in both the apo-enzyme (ZmC4-NADP-ME) and the secondary/ternary complex with NADP/pyruvate (SbC4-NADP-ME) is surprising because—based on previous work on non-plant NADP-ME structures25,29,31—one would rather expect such a conformational change following substrate and/or co-factor binding. These non-plant NADP-ME structures uncovered an open/closed switch transition, with the closed form probably corresponding to the catalytically competent conformation while the open form is probably required for substrate binding and product release. This open–closed transition during catalysis has been posited as being general for malate enzymes31. According to our results, the 'open' and 'closed' conformations appear to be independent from ligand binding, which might hint at an alternative mode of activity.

Identification of amino acid residues involved in C_4 optimization

To identify amino acid residues that contribute to the observed C4-specific function of C4-NADP-ME isoforms, we searched for positions in the protein sequence alignment at which all C4 isoforms share the same amino acid, while all non-photosynthetic isoforms from C4 (nonC4) or C3 plants share a different amino acid. We first focused on the single C4 lineage Andropogoneae within subfamily Panicoideae of the Poaceae, which includes the C4 grasses maize and sorghum. As a C3 reference species, we also included Oryza sativa (rice), another member of the Poaceae. Based on our alignment, we identified 20 amino acids as being differentially conserved between the isoforms analysed (Supplementary Fig. 5).

To focus on the most promising of these 20 candidates for further functional analysis, we used the crystal structures of ZmC4-NADP-ME and SbC4-NADP-ME. Based on analysis of the structures, we identified the subset of sites that (1) could affect substrate/co-factor binding or catalysis, (2) might be involved in dimer or tetramer formation or (3) might be part of a putative allosteric malate inhibition site. These candidates were F140, N142, N159, T163, R201, E339, P433, Q503 and L544 (Fig. 2).

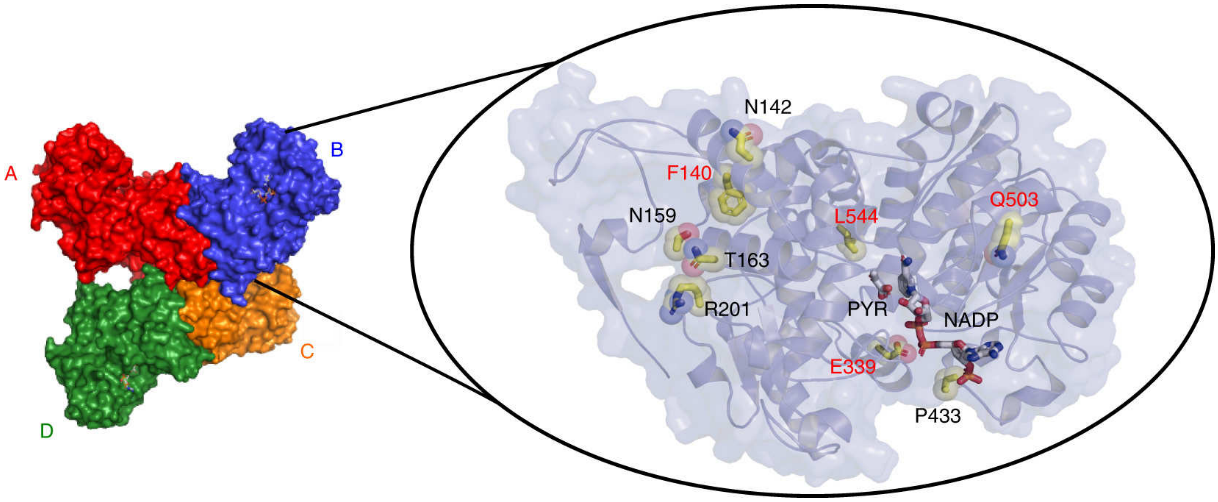
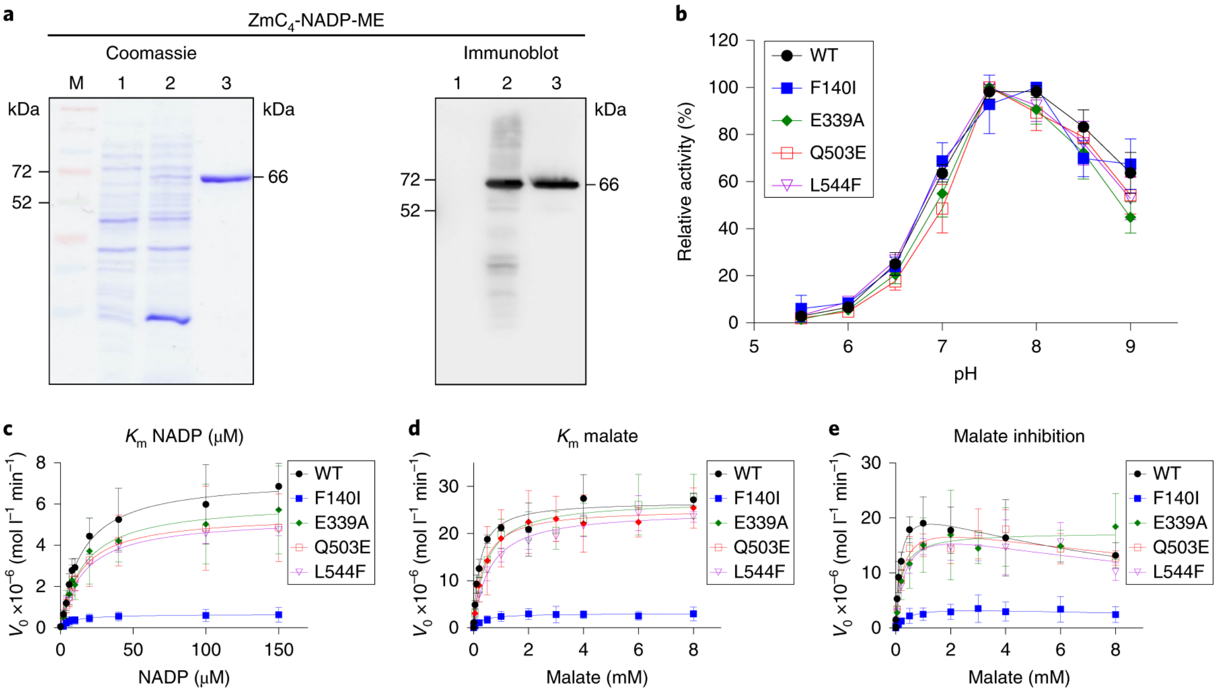


Fig. 2: Candidate amino acids for functional analysis identified using Zm- and SbC$_4$-NADP-ME crystal structures. The C$_4$-NADP-ME tetramer is shown on the left (colour coding as in Fig. 1a). A detailed view of monomer B is seen on the right, showing all candidate amino acids for functional analyses. The four amino acids are also strictly differentially conserved in *S. italica* and the examined substitution mutants in this study are shown in red. Refer to Supplementary Video 1 for a 3D overview of the position of the four mutated amino acids. F140, N142, N159 and T163 represent positions comparable to the fumarate binding site (allosteric activator) in human malic enzyme (PBD 1PJ3); R201 is in close proximity to N159 and points towards the dimer interface; E339 is located in helix αB4, which interacts with NADP; the rigid P433 is close to K435, which binds to NADP; Q503 is part of helix αC5, and L544 is part of helix αB5. The position of NADP and pyruvate (PYR) indicates the active site.

To further restrict the number of amino acids identified in the lineage Andropogoneae to the most promising candidates, we extended our analysis to *Setaria italica*, another member of the family Panicoideae of the Poaceae that belongs to an independent C$_4$ linage in the Paniceae[10,32]. We reasoned that *Setaria*'s unique plastidic NADP-ME sequence (SiC$_4$-NADP-ME) has also been optimized for its role in C$_4$ photosynthesis. We found that only four of the amino acids of putative functional importance are strictly differentially conserved in the extended alignment (F140, E339, Q503 and L544 in ZmC$_4$-NADP-ME; Fig. 2 and Supplementary Video 1). All amino acids identified in C$_4$-NADP-ME originated through single point mutations of the respective codons in the nonC$_4$-isofom. We thus decided to focus on these four sites for mutagenesis of ZmC$_4$-NADP-ME, and for further biochemical and structural characterization.

We produced mutant variants of ZmC$_4$-NADP-ME in which we independently substituted the four amino acids for those conserved in the nonC$_4$-NADP-MEs: F140I, E339A, Q503E and L544F. The recombinantly expressed and purified proteins showed the expected molecular mass (66 kDa; Fig. 3a and Supplementary Fig. 6) and presented similar dependencies of the activity on the pH of the reaction media, with an optimal pH of 7.5–8.0 (Fig. 3b).



f

ZmC_4-NADP-ME	WT	F140I	E339A	Q503E	L544F
K_m malate (mM)	0.23 ± 0.02	0.47 ± 0.13*	0.47 ± 0.17*	0.46 ± 0.12*	0.58 ± 0.04**
K_m NADP (μM)	14.9 ± 2.2	8.2 ± 2.8*	15.9 ± 2.2	12.8 ± 3.4	15.9 ± 2.8
k_{cat} (s^{-1})	28.1 ± 5.1	3.6 ± 1.3****	28.7 ± 3.3	29.8 ± 5.5	27.4 ± 1.8
k_{cat}/K_m (s^{-1} mM^{-1})	122.2	7.7	61.1	64.8	47.2
K_i malate$^{\#}$ (mM)	9.1 ± 2.6	15.8 ± 6.6	–	16.7 ± 2.4	13.4 ± 8.0

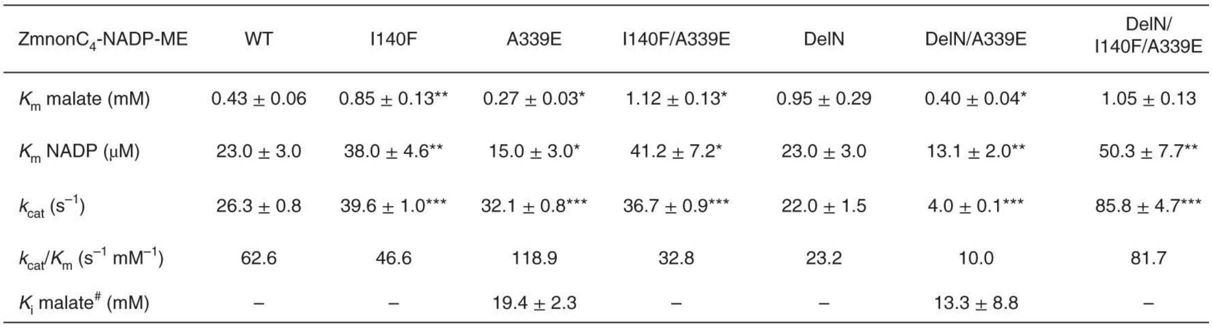
Fig. 3: Comparison of ZmC_4-NADP-ME variants shows that the identified amino acids contribute to the kinetic differences between C_4- and nonC_4 isoforms. a, SDS–PAGE stained with Coomassie (left) and analysed by immunoblot using anti-His-HRP conjugate antibodies (right) of different protein fractions during the isolation of recombinant wild-type (WT) ZmC_4-NADP-ME. Lanes: M, molecular weight markers, Spectra Multicolor Broad Range Protein Ladder (ThermoFisher Scientific); 1, 10 μg of non-induced cell culture lysate; 2, 10 μg of cell culture lysate 16 h after induction of protein production; 3, affinity-purified ZmC_4-NADP-ME, 1.5 μg for protein staining and 0.5 μg for immunoblot. The size of the recombinant ZmC_4-NADP-ME (66 kDa) is indicated on the right. The experiment was repeated three times independently with similar results. **b**, Dependence of the activity of ZmC_4-NADP-ME WT and the four engineered enzyme versions on the pH of the reaction medium. For each enzyme version, the highest activity measured (reaction velocity (V_0), in μmol mg^{-1} min^{-1}: WT, 25.25; F140I, 1.42; E339A, 24.05; Q503E, 22.82; L544F, 21.30) was set to 100%. **c**, ZmC_4-NADP-ME activity under varying amounts of NADP at pH 8.0 using 50 ng enzyme per assay. **d**, ZmC_4-NADP-ME activity under varying amounts of malate at pH 8.0 using 200 ng enzyme per assay. **e**, ZmC_4-NADP-ME activity under varying amounts of malate at pH 7.0 using 200 ng enzyme per assay. **f**, Kinetic parameters of recombinant ZmC_4-NADP-ME WT and the four mutated versions. Data were adjusted to the Michaelis–Menten equation by non-linear regression with Prism 6 (GraphPad Software). The parameters were determined at pH 8.0 with the exception of K_i malate, which was determined at pH 7.0 ($^{\#}$). For E339A, no inhibition by malate was measured. Asterisks indicate that

the value is significantly different from that of the corresponding WT as evaluated by two-tailed *t*-test. **P* < 0.05; ***P* < 0.01; *****P* < 0.0001 (the precise *P* values are shown in Supplementary Table 3). **a**–**f**, Values represent mean ± standard deviation; *n* = three independent enzyme preparations, each measured in triplicate.

Determination of the catalytic rate (kcat) at pH 8.0 indicated that mutants E339A, Q503E and L544F behave like the WT enzyme (kcat = 28.1 ± 5.1), while F140I presented an eightfold reduced value (kcat = 3.6 ± 1.3) (Fig. 3c–f). We observed no significant differences in affinity for the co-factor NADP for E339A, Q503E and L544F compared to the WT (Fig. 3c,f), while F140I showed a higher affinity (Fig. 3c,f). Affinity for the substrate malate was decreased by approximately twofold in all mutants (Fig. 3d,f); the apparent Michaelis constant (Km) values were similar to that of the maize nonC4-NADP-ME (Km malate, 0.43 mM8,15).

We studied the inhibitory effect of malate on activity at pH 7.0 and found that the E339A substitution provoked a loss of this regulatory property (Fig. 3e,f). The E339A variant thus behaves like the nonC4-isoforms. The variants F140I, Q503E and L544F showed inhibition by malate at pH 7.0, with values of the inhibitor constant (Ki) similar to that of the WT ZmC4-NADP-ME (Fig. 3e,f). To further analyse the role of E339 in inhibition by malate we introduced this residue into ZmnonC4-NADP-ME, producing the mutant A339E. The kinetic characterization of the recombinant mutant at pH 8.0 indicated that the affinity for NADP and malate and the catalytic rate resemble those of ZmC4-NADP-ME (Fig. 4a). At pH 7.0, the A339E variant incurred malate inhibition with a Ki value twofold higher than that of ZmC4-NADP-ME (Fig. 4a). These results confirm a pivotal role of the evolutionary A339E substitution in the acquisition of malate inhibition by the C4 isoform, while also increasing kcat and decreasing Km for both co-factor and substrate.

a



ZmnonC$_4$-NADP-ME	WT	I140F	A339E	I140F/A339E	DelN	DelN/A339E	DelN/ I140F/A339E
K_m malate (mM)	0.43 ± 0.06	0.85 ± 0.13**	0.27 ± 0.03*	1.12 ± 0.13*	0.95 ± 0.29	0.40 ± 0.04*	1.05 ± 0.13
K_m NADP (μM)	23.0 ± 3.0	38.0 ± 4.6**	15.0 ± 3.0*	41.2 ± 7.2*	23.0 ± 3.0	13.1 ± 2.0**	50.3 ± 7.7**
k_{cat} (s^{-1})	26.3 ± 0.8	39.6 ± 1.0***	32.1 ± 0.8***	36.7 ± 0.9***	22.0 ± 1.5	4.0 ± 0.1***	85.8 ± 4.7***
k_{cat}/K_m (s^{-1} mM^{-1})	62.6	46.6	118.9	32.8	23.2	10.0	81.7
K_i malate$^{\#}$ (mM)	–	–	19.4 ± 2.3	–	–	13.3 ± 8.8	–

b

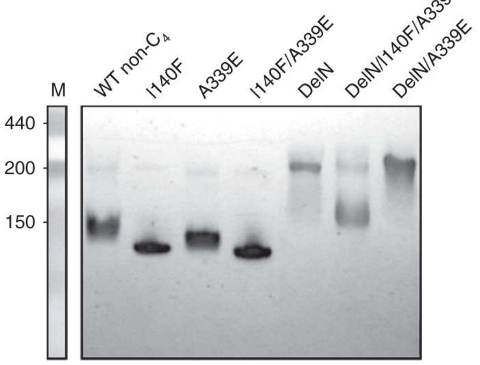


Fig. 4: Comparison of ZmnonC$_4$-NADP-ME variants confirms pivotal roles for substitutions I140F and A339E in combination with an N-terminal deletion. a, Kinetic parameters of recombinant

ZmnonC_4-NADP-ME WT and mutant variants A339E, I140F, I140F/A339E, DelN, DelN/A339E and DelN/I140F/A339E. Data were adjusted to the Michaelis–Menten equation by non-linear regression with Prism 6 (GraphPad Software). The parameters were determined at pH 8.0 with the exception of K_i for malate, which was determined at pH 7.0 ($^{\#}$). A dash (–) indicates that no inhibition by malate was determined. The values represent mean ± standard deviation; n = three independent enzyme preparations, each measured in triplicate. For statistical analysis, A339E, I140F and I140F/A339E were compared to WT, and DelN/A339E and DelN/I140F/A339E to DelN. Asterisks indicate that the value is statistically significantly different from the corresponding reference value as evaluated by two-tailed t-test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$ (the precise P values are shown in Supplementary Table 4). **b**, Native PAGE of recombinant ZmnonC_4-NADP-ME (WT nonC_4) and mutants I140F, A339E, I140F/A339E, DelN, DelN/A339E and DelN/I140F/A339E (20 mIU of each enzyme were loaded in lanes 1–4 and 10 mIU in lanes 5–7), followed by in-gel NADP-ME activity assay. Coomassie-stained native molecular weight markers (MWGF1000-1KT (Sigma-Aldrich)) are shown on the left (M). The experiment was repeated three times independently, with similar results.

Oligomeric organization of ZmC_4-NADP-ME variants

For analysis of whether the purified recombinant ZmC_4-NADP-ME folded correctly and whether any of the single mutations introduced in the engineered enzymes affected their secondary conformations, all proteins were analysed by circular dichroism spectroscopy. The spectra thereby obtained indicated that the engineered proteins generally have conserved secondary structures (Supplementary Fig. 7).

Recombinant ZmC_4-NADP-ME and variants E339A, Q503E and L544F assembled as tetramers at pH 8.0 (Fig. 5a). The F140I variant was found to be a mixture of oligomeric states, from tetramer to monomer (Fig. 5a). Native polyacrylamide gel electrophoresis (PAGE) of the ZmC_4-NADP-ME variants assayed for NADP-ME activity showed that the WT enzyme and all mutants are active only as tetramers under assay conditions (pH 8.0) (Fig. 5b). This is consistent with the finding that only the tetrameric form of C_4-NADP-ME is active in sorghum[11]. In line with reduction of the tetramer fraction in the F140I mutant, we observed only a faint band showing NADP-ME activity in the in-gel assay compared to the other variants (Fig. 5b), and k_{cat} eightfold lower relative to the WT (Fig. 3f). Together, these results indicate that the low k_{cat} of the F140I mutant is not due to lower kinetic performance of this enzyme but is related to the smaller fraction of the enzyme that is present with the active tetrameric structure.

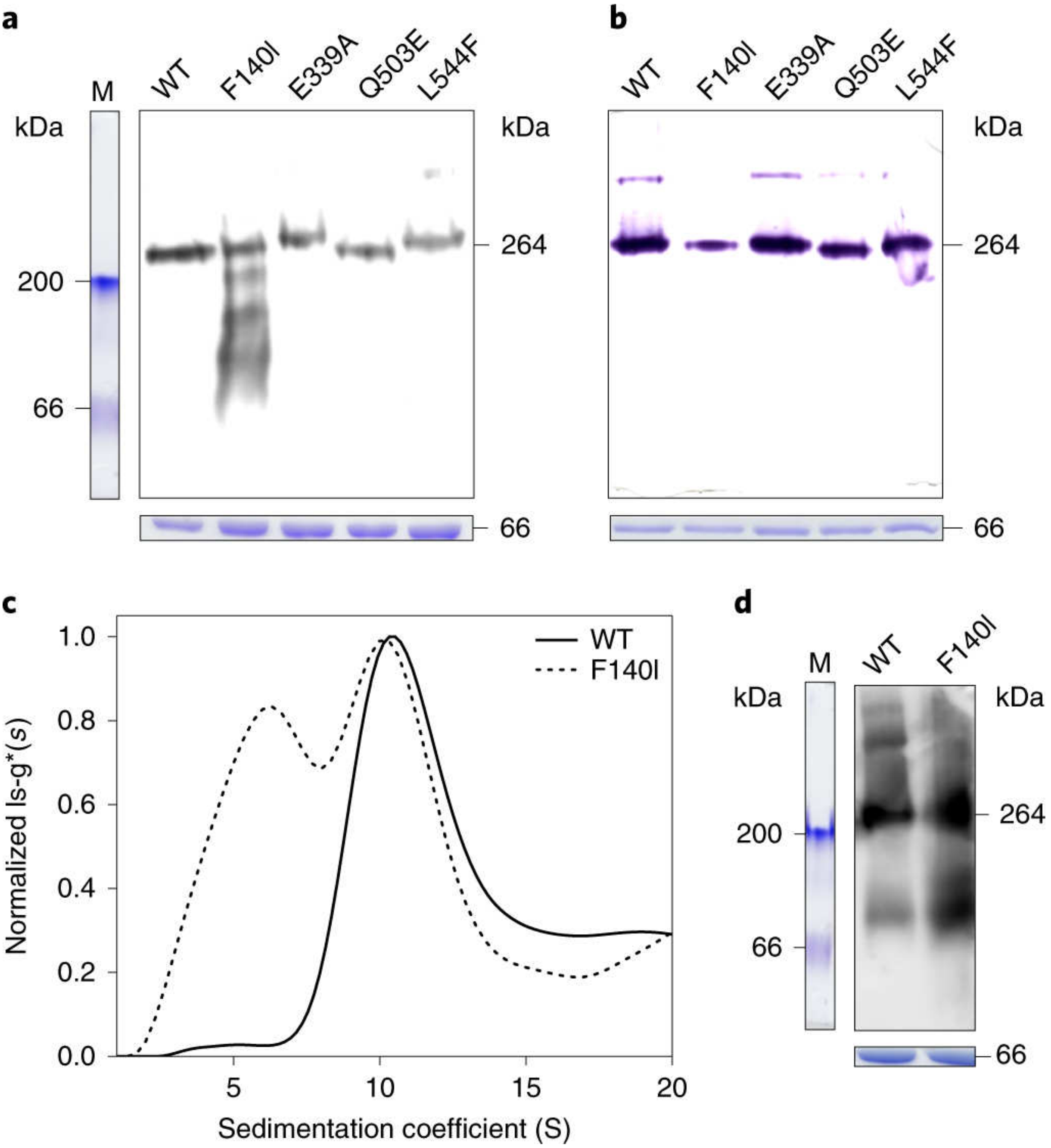


Fig. 5: F140, but none of the other strictly differentially conserved residues, contributes to stabilization of the tetramer. **a**, Native PAGE of recombinant ZmC$_4$-NADP-ME WT and the four mutated versions (3.0 µg protein per lane) followed by immunoblot analysis using anti-ZmC$_4$-NADP-ME antibodies. Coomassie-stained native molecular weight markers (bovine serum albumin (66 kDa) and ß-amylase (200 kDa;Sigma-Aldrich)) are shown on the left. **b**, Native PAGE of recombinant ZmC$_4$-NADP-ME WT and the mutated versions (1.0 µg protein per lane) followed by in-gel NADP-ME activity assay. **c**, Close-up view of the sedimentation coefficient distribution value 0–20 S of recombinant ZmC$_4$-NADP-ME WT and the F140I mutant at pH 8.0. Data were fitted using the ls-g*(*s*) model of the software package SEDFIT and normalized based on maximum peak height. **d**, Native PAGE of recombinant ZmC$_4$-NADP-ME WT and the F140I mutant (10 µg protein per lane) analysed by immunoblot using anti-ZmC$_4$-NADP-ME antibodies. The proteins were incubated at pH 7.0 for 16 h before electrophoresis. The expected molecular weight of the ZmC$_4$-NADP-ME homotetramer (264 kDa) is indicated on the right. Coomassie-stained loading controls (ZmC$_4$-NADP-ME, 66 kDa) are

shown below the native gels, and Coomassie-stained native molecular weight markers are shown on the left. The experiments in **a**,**b**,**d** were repeated three times independently, with similar results.

To further analyse protein quaternary organization, the WT enzyme and the F140I variant were subjected to analytical ultracentrifugation. The measurements thus acquired confirmed a stable tetrameric state of the WT ZmC_4-NADP-ME at pH 8.0, with a sedimentation coefficient of approximately 10 S (Fig. 5c). The F140I variant showed two broad peaks with sedimentation coefficients of approximately 5 and 10 S (Fig. 5c). The fraction of F140I found in these peaks is much lower compared to the WT (Fig. 5c shows normalized values). This is due to the presence of protein agglomerates that form under the conditions of the assay at sedimentation coefficients higher than 20 S. With F140I, 81% of the total protein was always obtained as agglomerates whereas WT ZmC_4-NADP-ME agglomerates represent only 29.6% of total protein.

We additionally examined the oligomerization state of the recombinant WT ZmC_4-NADP-ME after overnight incubation at pH 7.0 before separation in native PAGE. Detection of protein by immunoblot analysis indicated the presence of a low oligomeric form in the WT, with a gel mobility similar to that observed for the F140I variant treated under the same conditions (Fig. 5d).

To further analyse the role of F140 in the stabilization of the oligomeric structure of ZmC_4-NADP-ME, we introduced this amino acid into the ZmnonC_4-NADP-ME background, producing the I140F mutant. Native PAGE assayed for NADP-ME activity shows that I140F assembles as an active dimer but presents a slightly higher mobility than the WT ZmnonC_4-NADP-ME and A339E dimers (Fig. 4b), suggesting a different arrangement of its structure. The kinetic characterization of I140F at pH 8.0 indicated lower affinity for NADP and malate, and increased k_{cat} compared to the WT ZmnonC_4-NADP-ME (Fig. 4a). At pH 7.0, the I140F mutant was not inhibited by malate (Fig. 4a).

Use of combination of mutants to analyse the evolution of C_4 properties

To further analyse the evolution towards C_4-NADP-ME, we introduced the double mutation I140F/A339E into ZmnonC_4-NADP-ME (Fig. 4a). We found that the kinetic behaviour of the double mutant at pH 8.0 is comparable to that of the I140F mutant; also, I140F/A339E does not show malate inhibition at pH 7.0 (Fig. 4a). Interestingly, I140F/A339E presents the same mobility in native gels as I140F (Fig. 4a). These results indicate that the overall structural organization of I140F/A339E and I140F influences the parameters measured; moreover, in this structural background, the mutation A399E is not effective in conferring malate inhibition.

Besides the changes in amino acids differentially conserved among NADP-ME isoforms, they differ in the N-terminal region: nonC_4-NADP-MEs show 9–15 extra amino acids in domain A when compared to C_4-NADP-MEs (Supplementary Fig. 5). The N-terminal deletion of ZmnonC_4-NADP-ME (DelN, Supplementary Fig. 5) triggers the tetramerization of the enzyme[11] (Fig. 4b). Thus, to further analyse the role of F140 and E339 in conferring C_4 properties to NADP-ME, we introduced these C_4 amino acids into DelN, generating the mutants DelN/A339E and DelN/I140F/A339E.

DelN/A339E has the same mobility in native PAGE as DelN, indicating that is able to tetramerize (Fig. 4b). Compared to DelN, the mutant DelN/A339E presents higher affinity for malate and NADP and a drastically decreased k_{cat} at pH 8.0. The mutation A339E introduces inhibition by malate at pH 7.0

also in the DelN background (Fig. 4a). Interestingly, DelN/I140F/A339E presents higher mobility in native PAGE than DelN and DelN/A339E, recalling the induction of a change in mobility observed by the introduction of I140F in both WT ZmnonC$_4$-NADP-ME and the mutant A339E (Fig. 4b). Compared to DelN, DelN/I140F/A339E has increased K_m for NADP and k_{cat} at pH 8.0; the k_{cat} value measured is 3.9-fold higher than that of DelN/A339E, 21-fold higher than that of DelN and 3.3-fold higher than that of the WT, resulting in 31% higher catalytic efficiency compared to WT ZmnonC$_4$-NADP-ME (Fig. 4a). Again, we observed that the mutation A399E is not effective in introducing malate inhibition at pH 7.0 in a ZmnonC$_4$-NADP-ME variant with altered overall structure due to the I140F mutation (Fig. 4a).

Discussion

Our results suggest that amino acid 140 is in a position critical for the stabilization of the dimeric (nonC$_4$) or tetrameric (DelN and C$_4$) states of NADP-ME. The introduction of the mutation I140F in nonC$_4$-NADP-ME WT and DelN changes the mobility of these mutants in native gels (Figs. 4b and 5a); this modification of the overall structure of the proteins might explain the changes observed in the kinetic parameters. Moreover, the mutation F140I in ZmC$_4$-NADP-ME destabilizes the oligomeric organization of the protein, producing dimers and even monomers (Fig. 5a). In C$_4$-NADP-ME, F140 is near the N terminus, which is located close to the dimer interface (Supplementary Video 2). F140 is part of helix αA3, which is involved in monomer–monomer interactions at the dimer interface. While F140 does not interact directly with the neighbouring monomer (Fig. 2 and Supplementary Table 2), the adjacent amino acids K138 and K139 form hydrogen bonds with E288 and G196 of the adjacent protein chain (Fig. 6a and Supplementary Video 2). We therefore speculate that replacement of F140 by I140 disturbs the helical arrangement of that structural element (αA3), which in turn precludes the hydrogen bonding pattern of these two lysine residues, reducing dimer stability. In ZmC$_4$-NADP-ME, F140 may facilitate the generation of tetramers by stabilizing the dimers.

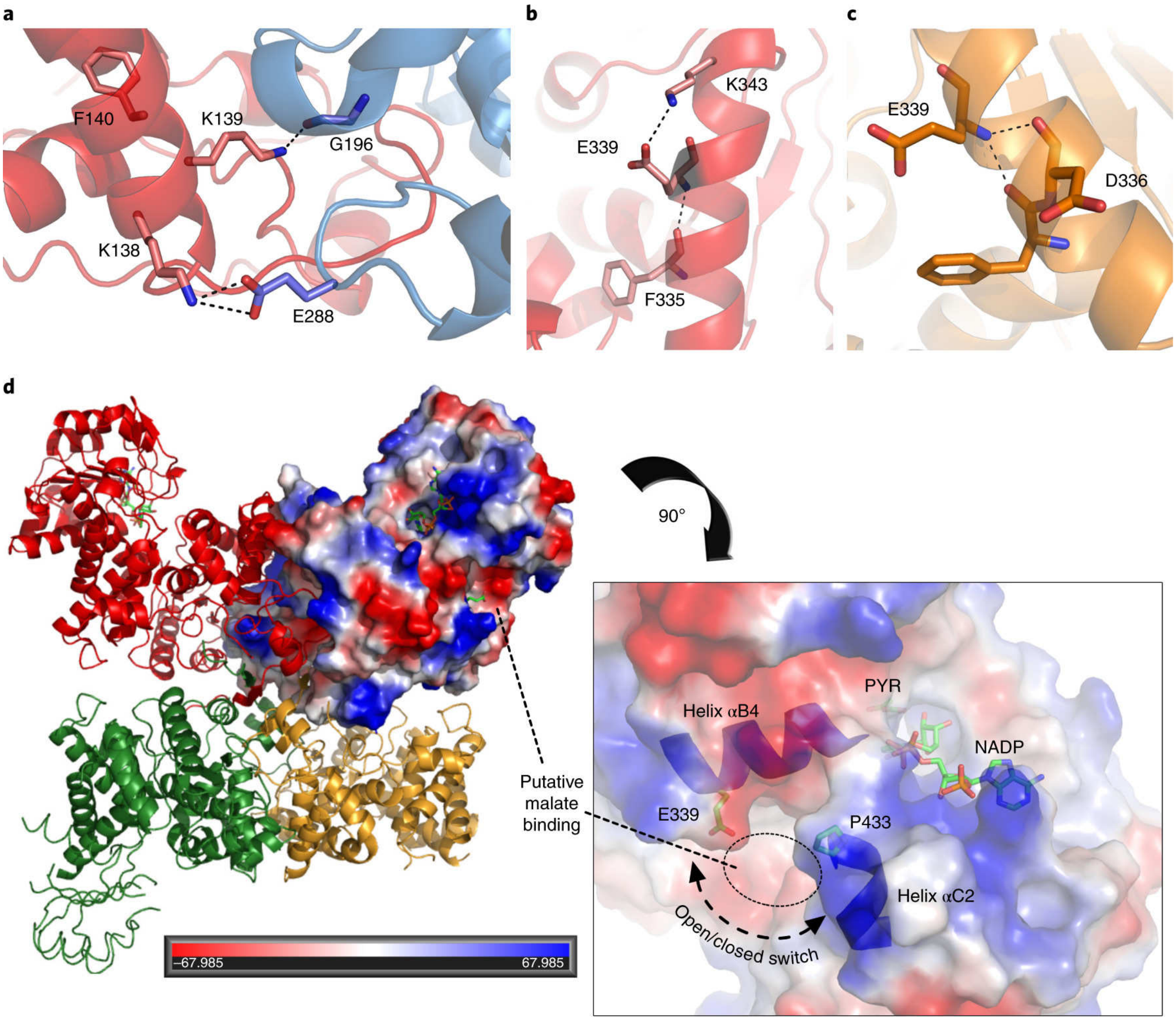


Fig. 6: Positions and interactions of F140 and E339 in the C_4-NADP-ME structure. a, In monomer A, K139 and K138 next to F140 are involved in dimer interactions with G196 and E288 from the adjacent monomer B in ZmC_4-NADP-ME. Refer to Supplementary Video 2 for a 3D visualization of the position of F140. **b**, E339 in all four monomers interacts with F335 and, in monomer A, along with K343 in Zm-C_4NADP-ME. **c**, E339 in monomer D interacts with D336 in Zm-C_4NADP-ME. Red, monomer A; blue, monomer B; orange, monomer D. **d**, Electrostatic surface representation of monomer B in the SbC_4-NADP-ME tetramer (left) and magnification of the co-factor binding site (right). Highlighted are the bound NADP molecule (coloured sticks), the position of E339, which allows delimitation of the malate allosteric binding site, and helix αB4. Red, monomer A; green, monomer C; orange, monomer D. The surface colouring of monomer B represents negative charge (red) to positive (blue).

We propose that the evolutionary acquisition of F140 by C_4-NADP-ME was necessary to assist in stabilization of the oligomeric assembly of the protein, which changes with stromal pH depending on the time of day/night[13]. Previous work conducted in vitro on sugar cane C_4-NADP-ME showed that a change in pH from 8.0 to 7.0 provoked a dissociation of the tetrameric enzyme into a dimer[13]. Here, we show that after lengthy incubation of ZmC_4-NADP-ME at pH 7.0, part of the enzyme is also found in an inactive, low-oligomeric state (Fig. 5d). Our results indicate that during the night, when the C_4 pathway is not active and stromal pH decreases to 7.0, C_4-NADP-ME loses its quaternary structure and adopts an inactive oligomerization state (dimer or monomer). These changes in oligomerization may represent a means to modulate enzymatic activity dependent on metabolic demand. The use of a higher oligomeric organization for enzyme regulation was suggested for the human mitochondrial NAD(P)-ME; dissociation of the tetramer into dimers causes this enzyme to be less active, less suitable for malate binding and less responsive to its allosteric regulator[33].

The protein region that extends from amino acid residue 248 to the C terminus of C_4-NADP-ME is involved in the inhibition of malate at acidic pH[15]. Lack of inhibition of the C_4-NADP-ME at pH 8.0 suggests that the amino acid residue(s) involved in the allosteric regulation change their protonation state(s) depending on the surrounding pH. We found that E339 of ZmC_4-NADP-ME is directly involved in the inhibition of malate at pH 7.0. In fact, the nonC_4-isoforms possess a hydrophobic A at this position while the C_4 isoforms possess E, a polar and acidic amino acid. We showed that substitution E339A in ZmC_4-NADP-ME is sufficient to completely eliminate the pH-dependent regulation of this isoform (Fig. 3) and vice versa—that the change from A to E in nonC_4-NADP-ME is sufficient to induce malate inhibition (Fig. 4).

The crystal structures of Zm- and SbC_4-NADP-ME show that E339 is far from the catalytic site and is located on the surface of helix αB4, with its side chain oriented outwards to the solvent (Fig. 2 and Supplementary Video 3). The backbone and side chain of E339 are involved only in intramolecular hydrogen bonding with neighbouring residues on the same helix (in all chains to F335, in chain A additionally to K343 and in chain D to D336; Fig. 6b,c). The structural connections between E339 and the active site suggest three possible molecular mechanisms for the observed pH-dependent malate inhibition. First, the N-terminal tip of helix αB4 is directly engaged within the active site (Fig. 6 and Supplementary Video 3). Hence, after malate binding at the putative allosteric site (Fig. 6d and Supplementary Video 3), E339 could induce a positional rearrangement of the helix, affecting substrate binding. Second, E339 is well positioned to interact with helix αC2 (broken in two by P433) and is thus ideally placed to modulate the open/closed transition essential for enzymatic activity[31] (Fig. 6d and Supplementary Video 3). The distance between the alpha carbons of E339 and P433 changes from 12 to 15.6 Å in the transition from closed to open conformation in SbC_4-NADP-ME. Binding of malate in the vicinity of E339 could hamper open/closed switching, with consequent catalysis inhibition (Fig. 6d and Supplementary Video 3). Finally, in consideration of the distribution of negative electrostatic potential observed in the putative allosteric site (Fig. 6d), it cannot be ruled out that malate actually does not bind in this region and that E339 acts as a pH sensor; protonation at pH 7.0 may induce a conformational change in the active site of the enzyme via helix αB4 (Fig. 6 and Supplementary Video 3). This rearrangement might induce the binding of malate at the active site in a less effective manner, rendering the observed malate inhibition. Independent of the underlying molecular mechanism, it is clear that during the night, when stromal pH decreases to 7.0–7.5, E339 enables the inhibition by malate of C_4-NADP-ME, and that this inhibition does not take place during the day when photosynthesis is active and stromal pH rises to 8.0. The lack of malate inhibition at pH 8.0 then facilitates a high flux of malate decarboxylation to provide CO_2 for photosynthetic fixation; such flux during the night would simply produce carbon loss. In a similar way, the activity of C_4-PEPC is also inhibited by malate[34]; light-dependent phosphorylation of C_4-PEPC releases this inhibition, rendering a more active enzyme when photosynthetic flux is high.

The amino acid changes independently introduced by Q503E and L544F in ZmC_4-NADP-ME produced enzyme variants with reduced affinity for the substrate malate. The K_m values measured were around twofold higher than that of the WT ZmC_4-NADP-ME, and thus resemble that of the nonC_4-isoform (K_m

malate, 0.43 mM[8]). These results show that the residues identified represent adaptive changes that gave rise to the increased affinity of C_4-NADP-ME for its substrate. During neo-functionalization of the chloroplastic NADP-ME, E503 of the ancestral isoform changed to Q, a (polar) uncharged amino acid (Fig. 2). Crystal structures indicate that Q503 is positioned on helix αC5 on the protein surface (Fig. 2), forming intramolecular interactions with adjacent amino acids (Supplementary Fig. 8a,b). The other adaptive change involved the introduction of the aliphatic L at position 544, which was occupied by the aromatic F in the ancestral enzyme. The crystal structures show that L544 is part of helix αB5 buried within the protein, and forms intramolecular interactions with adjacent amino acids (Supplementary Fig. 8c).

The adaptive changes discussed so far were inferred based on the analysis of single amino acid substitutions in maize C_4-NADP-ME and nonC_4-NADP-ME. The characterization of combinations of mutations provided a more comprehensive view of the molecular adaptations of C_4-NADP-ME. The pH-dependent malate inhibition gained through the substitution A339E in both WT and DelN nonC_4-NADP-ME was lost when adding the substitution I140F (I140F/A339E and DelN/I140F/A339E, respectively), indicating that destabilization of the oligomeric assembly induced by F140 in the nonC_4-NADP-ME background also impacts negatively on malate inhibition at pH 7.0. This possibly occurs through alteration in the connection of E339 with the active site caused by the conformation changes induced by the I140F mutation, independently of the oligomerization state. Interestingly, the introduction of E339 in the tetrameric DelN mutant adds another C_4 feature, a higher substrate affinity; however, a penalty in the catalytic rate is observed. In turn, this penalty in catalytic rate is overcome when F140 is introduced although this change generates a decrease in affinity for NADP and the loss of inhibition by malate at pH 7.0. Thus, mutational effects are dependent on the presence of other mutations—that is, we observed substantial epistasis between individual mutations. This may not be surprising, as biochemical changes induced by the substitutions appear to rely predominantly on long-range effects mediated by changes in the structural details of the protein (Fig. 4). It is likely that further substitutions compensate for the negative effects of I140F in the DelN/A339E background, for example by further stabilizing the protein oligomeric structure.

In sum, our work defines molecular adaptions that occurred during the evolution of C_4-NADP-ME from a housekeeping isoform. We found that while Q503, L544 and E339 independently confer high affinity for malate, E339 is sufficient to confer both high catalytic efficiency and inhibition by malate at pH 7.0. In addition, F140 is critical for the stabilization of the oligomeric structure of C_4-NADP-ME while the N-terminal region is responsible for tetramerization. These results provide tools for initiation of analyses of the physiological significance of enzyme regulatory mechanisms.

Our results provide molecular understanding that can guide the creation of enzyme versions suitable for rational synthetic biology approaches. For example, to enhance plant yield by installing the C_4 trait into a C_3 plant[35,36,37,38] it is important not only to guarantee high and localized expression of C_4 enzymes, but also to fine-tune their activities to facilitate the required photosynthetic flux through the C_4 cycle during the day while avoiding carbon loss during the night. We show that, at least for NADP-ME, this important goal may be achieved by introducing a small number of mutations into a copy of the C_3 plant's own enzyme rather than by introducing a foreign C_4 isoform.

Methods

Sequence selection and identification of C$_4$-specific amino acids

We obtained the amino acid sequences of the plastidic C$_4$- and nonC$_4$-NADP-ME isoforms of *Z. mays* cultivar Ensembl-18 (sequence IDs: GRMZM2G085019_T01, GRMZM2G122479_P01) and *S. bicolor* cultivar v.3.1.1 (sequence IDs: Sobic.003G036200.1, Sobic.009G108700.1) from the Phytozome 11 database[39]. We used the maize C$_4$ and nonC$_4$-NADP-ME sequences as BLAST$^+$ (v.2.6.0) queries to identify the corresponding orthologues in the *S. italica* cultivar v.2.2 proteome in the same database (sequence ID: Seita.5G134300.1). We also used the maize nonC4-NADP-ME sequence as a BLAST$^+$ (v.2.6.0) query to identify the unique plastidial NADP-ME in the *O. sativa* cultivar v.7_JGI proteome (LOC_Os01g09320.1).

All sequences were aligned using PRANK (v.140603)[40]. We used the alignment of C$_4$- and nonC$_4$-NADP-ME isoforms (Supplementary Fig. 5) to identify amino acid positions that are strictly differentially conserved in C$_4$ compared to nonC$_4$ sequences: we selected all positions for which the C$_4$ sequences of *Z. mays* and *S. bicolor* contained the identical amino acid while a second, different, amino acid was found in both nonC$_4$ sequences of *Z. mays* and *S. bicolor* and the unique plastidial *O. sativa* sequence. To identify the subset of the most promising candidates of strictly differentially conserved amino acid positions, we repeated this analysis; next, we additionally required that the corresponding amino acid of the unique plastidial NADP-ME sequence of *S. italica* was identical to that of the C$_4$-NADP-ME of *Z. mays* and *S. bicolor*. Such a pattern of strict differential conservation would indicate that the respective amino acid position is functionally important in all isoforms, while the selected amino acid—and hence the molecular function—differs between C$_4$ and nonC$_4$/C$_3$ isoforms.

Generation of expression constructs and site-directed mutagenesis of ZmC$_4$-NADP-ME and ZmnonC$_4$-NADP-ME

To express ZmC$_4$-NADP-ME, the complementary DNA sequence coding the mature protein was amplified from the plasmid pET32*::*ZmC$_4$ NADP-ME[14] and cloned in the pET16b expression vector (Novagen) using Gibson assembly[41] and the following primer pairs: ‘Gib_FW’ (5′-GCCATCATCATCATCATCATCATCATCATCACAGCAGCGGCCATATCGAAGGTCGTCATGCGATGGTCTC-3′) and ‘Gib_RV’ (5′-GCAGCCAACTCAGCTTC CTTTCGGGCTTTGTTAGCAGCCGCTACCGGTAGTTGCGGTAGAC-3′). The resulting expression construct, *pET16b::ZmC$_4$-NADP-ME*, was used to introduce the desired point mutations by site-directed mutagenesis. Amplification of the whole expression construct was conducted in a PCR assay using the Physion Hot Start II DNA Polymerase (Thermo Scientific) and primer pairs in back-to-back orientation. Five identical 50-µl PCR reactions were performed simultaneously. The DNA obtained was mixed and subjected to self-ligation by incubation at 22 °C for 1 h in the presence of T4 DNA Ligase (Thermo Scientific). Further steps involved inactivation of the ligase by incubation of the reaction mix at 65 °C for 10 min, destruction of the still-present *pET16b::ZmC$_4$-NADP-ME* by digestion with DpnI (Thermo Scientific) at 37 °C for 2 h and DpnI inactivation by incubation at 80 °C for 5 min. For introduction of the F140I substitution, the primer pairs ‘F140I_FW’ (5′-[Phos]-TCATGAACACCCTGCG-3′ and ‘F140I_RV’ (5′-[Phos]TCTTCTTAATTTGGAGTTCCTG-3′) were used. The expression plasmid was called *pET16b::F140I_NADP-ME*. For the introduction of the E339A substitution, the primer pairs ‘E339A_FW’ (5′-[Phos]-AAAATATAGCAAGAGCCATC-3′) and ‘E339A_RV’ (5′[Phos]-GCAAGCAAATCAAAGGCATTGTG-3′) were used. The resulting expression plasmid was called *pET16b::E339A_NADP-ME*. For the introduction of the Q503E substitution, the primer pairs

'Q503_FW' (5'-[Phos]-AAGCATATACATGGAGT-3') and 'Q503_RV' (5'[Phos]-CTTCAGCAGTGCATTCA-3') were used. The expression plasmid was called *pET16b::Q503E_NADP-ME*. For introduction of the L544F substitution, the primer pairs 'L544F_FW' (5'-[Phos]-TCGGCCTCGGTCTTGTGATCTC-3') and 'L544F_RV' (5'[Phos]-ATCCAGGGAAAATGTAGGCATTG-3') were used. The expression plasmid was called *pET16b::L544F_NADP-ME*.

To express ZmnonC$_4$-NADP-ME, the cDNA sequence coding the mature protein was cloned in pET28 expression vector (Novagen) between the restriction sites *Nde*I and *Xho*I. First, a 1279-bp fragment was amplified from the plasmid *pET32::ZmnonC*$_4$ *NADP-ME*[14] using the following primer pairs: 'NdeImadI' (5'-CATATGGCCGCGGAGATGGAGCAG-3') and P4 (5'-GAAAATGTAGGCGTTGTTCGACTGT-3'). The fragment obtained by digestion of the PCR product with *Nde*I and *Bam*HI was ligated with a second fragment obtained by digestion of the plasmid *pET32::ZmnonC*$_4$ *NADP-ME* with *BamH*I and *Xho*I, and the vector pET28 digested with *Nde*I and *Xho*I. The resulting expression construct, *pET28::ZmnonC*$_4$*-NADP-ME*, was used to introduce the A399E mutation by site-directed mutagenesis essentially as described above, using the following primer pairs: 'FOR_nonC4A339E' (5'-AAACAAGATGGCTCTTGCTATATTTTTCAAGCAAATTAAAAGCATTATGATTC-3') and 'REV_nonC4A339E' (5'-GAATCATAATGCTTTTAATTTGCTTGAAAAATATAGCAAGAGCCATCTTGTTT-3'). The mutation I140F was obtained as described in ref. [14]. The oligonucleotide primer that introduced the mutation in the WT protein was 'I140F' (5'-CTCCAAATTAAGAAGTTCATGCACAACCT-3'), in which the mutation position is underlined in the oligonucleotide sequence. The fragment containing the mutation was subcloned from the PCR product to *pET32::ZmnonC*$_4$ *NADP-ME*, replacing the WT sequence. The N-terminal deletion of ZmnonC$_4$-NADP-ME (DelN) was previously obtained[11]. To introduce the mutation A339E in DelN (DelN/A339E), the C-terminal fragment of the plasmid *pET28::DelN_nonC4NADP-ME*, obtained by digestion with *EcoRI* and *XhoI* restriction enzymes, was replaced by the corresponding product obtained from the digestion of *pET28::A339E_nonC4NADP-ME*. The coding fragments of I140F/A339E and DelN/I140F/A339E were obtained by chemical synthesis (Twiss Bioscience) and cloned in pET28 plasmid.

Expression and purification of recombinant C$_4$- and nonC$_4$-NADP-ME variants

All expression constructs were transformed into *Escherichia coli* Rosetta2(DE3) (pLysSRARE2) (Merck) for ZmC$_4$-NADP-MEs, and in *E. coli* BL21(DE3) cells in the case of ZmnonC$_4$-NADP-MEs. After plasmid preparation, the success of the site-directed mutagenesis procedure was confirmed by sequencing. For heterologous protein production, transformed cells were grown in 400 ml lysogeny broth medium in the presence of the appropriate antibiotic in each case at 37 °C and 110 r.p.m., to an OD$_{600}$ of 0.6–0.8. To induce the expression of the heterologous protein, 1 mM isopropyl-ß-d-thiogalactopyranoside was added to the culture and the cells were incubated for 20 h under the same conditions. The cellular culture was harvested at 4.000*g* for 15 min, and pellets were stored at −20 °C until use.

For protein extraction, pellets were thawed on ice, resuspended in 20 mM Tris-HCl (pH 8.0) containing 500 mM NaCl, 5 mM imidazole, 2 mM phenylmethanesulfonyl fluoride and a spatula-tip amount of lysozyme, sonicated and centrifuged at 14,000*g* for 20 min at 4 °C. The supernatant was used for protein purification using gravity-flow immobilized metal ion chromatography on nickel-nitrilotriacetic acid agarose (Ni-NTA Agarose, Qiagen). The column was pre-equilibrated with 20 mM Tris-HCl buffer containing 500 mM NaCl and 5 mM imidazole. After loading of the supernatant, the columns were washed in four steps with 500 mM NaCl in 20 mM Tris-HCl (pH 8.0) containing increasing concentrations of imidazole (5, 50, 60 and 65 mM). Protein elution was performed in four steps of 500 µl of 20 mM Tris-HCl, 500 mM NaCl and 300 mM imidazole. The protein eluted in the first fraction was used for further kinetic measurements.

Removal of the His-tag for crystallization of Zm- and SbC$_4$-NADP-ME

For crystallization studies of ZmC$_4$-NADP-ME, all elution fractions were pooled and subjected to concentration and buffer exchange in a Vivaspin20 centrifugal concentrator (Sartorius AG) using 20 mM Tris-HCl (pH 8.0), 100 mM NaCl and 2 mM CaCl$_2$. Final concentrations of the protein samples ranged from 10 to 14 mg ml^{-1}. The N-terminal His-Tag of the recombinant enzyme was cleaved off by incubation with factor Xa protease (New England Biolabs), the recognition sequence for which was encoded by the pET16b vector and is located between the His-tag and the ZmC$_4$-NADP-ME sequences. Protease (1 µg) was added to 50 µg of ZmC$_4$-NADP-ME and incubated at 23 °C for 3 h. Cleaved His-Tag and factor Xa protease were eliminated by size exclusion chromatography using a HiPrep 16/60 Sephacryl S-300 High Resolution column (GE Healthcare Life Sciences), pre-equilibrated with 10 mM Tris-HCl (pH 8.0) containing 0.1 M NaCl. Gel filtration was carried out using the ÄKTA prime plus system (GE Healthcare Life Sciences). Elution was carried out using 120 ml of the same buffer at 0.2 ml min^{-1}. The protein eluted as a single peak at the volume expected for the homotetramer. In the case of SbC$_4$-NADP-ME, the His-tag fusion was removed by incubation with thrombin/SbC$_4$-NADP-ME (1/20) for 2 h at 16 °C. The cleaved His-tag was eliminated using a protocol similar to that previously described for ZmC$_4$-NADP-ME, by size exclusion chromatography. In each case, protein fractions were combined and subjected to concentration and buffer exchange using 10 mM Tris-HCl (pH 8.0) containing 5 mM MgCl$_2$ employing a Vivaspin20 centrifugal concentrator (Sartorius AG). For further crystallization studies, the protein concentration of the sample was adjusted to 10 mg ml^{-1} in the case of ZmC$_4$-NADP-ME and 6 mg ml^{-1} for SbC$_4$-NADP-ME.

Zm- and SbC$_4$-NADP-ME crystallization and X-ray data collection

The initial crystallization condition for ZmC$_4$-NADP-ME without His-tag was identified from sparse matrix screening using commercial screens (Molecular Dimension and Nextal, Qiagen). Crystals were grown at 12 °C by the sitting-drop vapour-diffusion method. Equal volumes (0.1 µl) of the enzyme (10 mg ml^{-1}) and reservoir solution were mixed and equilibrated against 40 µl reservoir solution. The most promising condition contained 0.2 M potassium chloride, 0.1 M sodium citrate (pH 5.5) and 37% pentaerythritol propoxylate (5/4 PO/OH) from MemGold I, B11 (Molecular Dimensions), and was optimized with protein at 5 mg ml^{-1} by a grid screen (the optimal diffracting crystals grew in 0.1 M potassium chloride, 0.1 M sodium citrate (pH 5.5) and 34% pentaerythritol propoxylate). Crystals appeared within a few days and grew to a size of about 50 × 50 × 250 µm^3 in 5–7 days. The crystals were fished after overlaying the drop with 1 µl mineral oil and flash-frozen in liquid nitrogen for X-ray data collection at 100 K. X-ray diffraction data were collected at a wavelength resolution of 0.9763–2.2 Å with beamline ID23-1 at ESRF.

In the case of SbC$_4$-NADP-ME, initial identification of crystallization conditions was carried out using the sitting-drop vapour-diffusion method with a robotic workstation (Honeybee963, Isogen Life Science). Sitting drops were set using 400 nl of a 1:1 mixture of protein and mother liquor, equilibrating against 150 µl mother liquors in the reservoir of Greiner plates, at both 4 and 20 °C. The mother liquors screened were obtained from commercial kits, Crystal screens HR2-110 and HR-112

(Hampton Research) and PEG Suite solutions (Qiagen). The best crystal hit contained 0.2 M KF and 20% PEG 3350 as mother liquor. The crystal growth habit comparing both crystallization temperatures was similar, so we continued working only at 20 °C. Initial SbC_4-NADP-ME crystals were manually optimized by varying precipitant and protein concentrations using VDX plates (Hampton Research) with a hanging-drop set-up. The best crystallization condition was ultimately achieved using 0.2 M KF and 30% PEG 3350 reservoir solution at 20 °C, with protein at 5.0 mg ml^{-1} in 20 mM Tris pH 8.0, 50 mM NaCl, 2 mM NADP, 40 mM pyruvate and 10 mM $MgCl_2$. Crystals appeared in about 7 days. Cryo-protection was achieved by slowly adding cryo-protection solution (25% (v/v) glycerol, 30% (w/v) PEG 3350, 0.1 M Tris-HCl (pH 8.0), 0.2 M $MgCl_2$, 2 mM NADP and 40 mM pyruvate) to the drop in small volumes (~5% drop volume step-wise, to >15% glycerol), followed by rapid soaking in 100% cryo-protection solution and flash-cooling in liquid N_2 for storage until data collection. X-ray diffraction data collection was performed with a Micromax007-HF generator (Rigaku) at 100 K with a rotating Cu anode at a wavelength of 1.5418 Å, Varimax-HF (Rigaku) optics and a Mar345 image plate detector (Mar Research) reaching 2.0 Å resolution (Extended Data Table).

Data processing, structure determination and refinement

X-ray diffraction data from ZmC_4-NADP-ME and SbC_4-NADP-ME were processed with XDS (v.November 11, 2013)[42] or Mosflm (v.7.1.0)[43] and scaled with Aimless (v.0.5.1)[44]. Phasing was performed by magnetic resonance using chain A from the human mitochondrial malic enzyme (PBD 1QR6) as search probe. The resulting structures were used as the starting point for further model building. Iterative cycles of refinement using Refmac5 (v.5.8.0238)[45] or Buster (v.2.10.3)[46] were performed after model rebuilding and validation using Coot[47]. After final refinement of ZmC_4-NADP-ME, 98.1% of all amino acids were in the favoured regions of the Ramachandran plot, 1.7% in the allowed regions and 0.2% were outliers. For SbC_4-NADP-ME, strong signs of flexibility were evident on monomers C and D. We thus excluded segments 423–434 and 524–529 in monomer C, as well as segments 365–383, 399–533 and 572–582 in monomer D, from the final model. In addition, entire portions of domain C on monomer C (residues 368–535 and 573–582) were refined as alternative conformations on both main chain and side chains. For SbC_4-NADP-ME, we modelled the co-factor NADP in monomers A, B and C (NADP is not entirely visible in monomer D) and pyruvate in two slightly different orientations with different occupancies in the active site of monomer B. After final refinement, 98.2% of all amino acids were in the favoured regions of the Ramachandran plot and 1.8% in the allowed regions. RMSD values were calculated using Superpose from the CCP4 suite (v.7.0.071), and interface areas were determined using the PDBePISA online server (http://www.ebi.ac.uk/pdbe/pisa/) (v.1.52). All images of the models were prepared using MacPyMOL (v.2.1.0)[48].

Protein quantification and gel electrophoresis procedures

The protein concentration of purified enzymes and other protein-containing samples used for analysis was determined using the Pierce BCA Protein Assay Kit (ThermoFisher).

SDS–PAGE was performed in 12% (w/v) polyacrylamide gels according to ref. [49]. Proteins were visualized by staining with Coomassie Brilliant Blue or electroblotted onto a polyvinylidene difluoride (PVDF) membrane (Carl Roth) for subsequent immunological detection. In every case, the recombinantly expressed and purified proteins showed the expected molecular mass of 66 kDa

(63.4 kDa of the mature ZmC_4-NADP-ME plus 2.5 kDa encoded by the expression vector, and 66 kDa of mature ZmnonC_4-NADP-ME after cleavage of the His-Tag fusion). For native PAGE, affinity-purified enzymes were separated on non-reducing and non-denaturing 6% (w/v) polyacrylamide gel at 100 V. The separated proteins were analysed ether by in-gel NADP-ME activity or immunological detection.

In-gel NADP-ME activity assay and immunological detection

In-gel NADP-ME activity assays were performed by incubation of the gels in the dark at room temperature in a reaction buffer containing 50 mM Tris-HCl (pH 8.0), 50 mM l-malate, 10 mM $MgCl_2$, 0.5 mM NADP, 0.05% (w/v) nitro blue tetrazolium and 3 mM phenazinemethosulfate.

For immunoblotting assays, the separated proteins were transferred to a PVDF membrane (Roti-PVDF, Carl Roth). The membranes were incubated for at least 1 h with a 1:7,500 dilution of Anti-His-Tag antibody coupled to horseradish peroxidase (Anti-His-HRP, Miltenyi Biotec) or, alternatively, with a 1:50 dilution of a rabbit polyclonal antibody raised against an epitope of ZmC_4-NADP-ME located on the solvent-exposed surface of the protein (peptide VWLVDSKGLIVD,Eurogentec). After washing several times, the membranes were incubated with a 1:2,500 dilution of Goat Anti-Rabbit IgG Antibody HRP-conjugate (Merck). Proteins were visualized by chemiluminescence with Pierce ECL Western Blotting Substrate (Thermo Scientific) on a LAS-4000 Mini Luminescent Image Analyzer (GE Healthcare Life Sciences).

Determination of kinetic parameters

NADP-ME activity was determined using a Synergy HT Biotek Plate Reader system by measuring the formation of NADPH at 340 nm at 25 °C. The standard assay medium contained 0.5 mM NADP, 10 mM $MgCl_2$, 4 mM l-malate, 50 mM Tris-HCl buffer (pH 8.0) and 50–200 ng enzyme per well in a final volume of 200 µl.

Dependence of activity on the pH of the medium was determined with the standard assay medium using different buffer systems: 50 mM MES (pH 5.5–6.5), 50 mM Tricine-Mops (pH 7.0–7.5) and 50 mM Tris-HCl (pH 7.5–9.0).

The apparent Michaelis constant (K_m) for malate and NADP was determined at pH 8.0 by varying the concentration of one component (NADP concentrations varied between 2 and 150 µM and malate concentrations between 0.001 and 8 mM for C_4-NADP-ME, and between 0.001 and 25 mM for nonC_4-NADP-ME) while keeping the concentration of the other component constant and saturated. Malate inhibition analysis was performed at pH 7.0.

An extinction coefficient (ε) of 6.22 mM^{-1} cm^{-1} for NADPH at 340 nm was employed in the calculations. All kinetic parameters were calculated with at least three batches of independently purified proteins, with each calculation containing a triplicate determination. Data were adjusted to non-linear regression with Prism 6 (GraphPad Software). When substrate inhibition was observed, the data were adjusted to a two-site model equation[15].

Secondary structure analysis by circular dichroism

Circular dichroism analyses of all recombinant versions of ZmC$_4$-NADP-ME were performed with proteins in 20 mM NaP$_i$ (pH 8.0) and 5 mM MgCl$_2$. Amicon Ultra 0.5-ml centrifugal filters with a nominal molecular weight limit of 50 kDa (Merck) were used for buffer exchange procedures. Before circular dichroism measurement, protein concentration was determined at 280 nm in a Jasco V-650 spectrophotometer and was adjusted to 0.16 mg ml^{-1}. The protein concentration, c (in mg ml^{-1}), was calculated with the equation $c = \mathrm{OD}_{280} \times M_w/(\varepsilon \times d)$, in which OD$_{280}$ is the measured optical density at 280 nm, M_r is the molecular mass of the protein (in g mol^{-1}), $\varepsilon = 61{,}770\ \mathrm{M}^{-1} \times \mathrm{cm}^{-1}$ is the molar extinction coefficient of the protein at 280 nm and $d = 1$ cm is the cell path length.

Circular dichroism spectra were obtained at 20 °C with a Jasco J-715 spectropolarimeter using a cell of 0.1 cm path length and averaging five repetitive scans between 250 and 190 nm. Mean residue ellipticity ($[\Theta]$) was obtained by the equation $[\Theta] = \mathrm{MRW} \times \Theta/(10 \times c \times d)$, in which MRW (mean amino acid residue weight) was calculated as the relationship of protein M_r to the number of peptide bonds ($N - 1$, where N is the number of amino acids in the protein chain), d is the cell path (0.1 cm), Θ is the ellipticity measured in milli-degrees (machine units) and c is the protein concentration obtained from the previous equation. The protein secondary structural composition was estimated from the experimental circular dichroism data using the CONTIN-LL method[50,51] at the Dichroweb online server[52] (http://dichroweb.cryst.bbk.ac.uk)[53]. Prism Software (http://www.graphpad.com/scientific-software/prism/) was used for visualization of the deconvolution by Dichroweb.

Quaternary structure analysis by analytical ultracentrifugation

Sedimentation velocity experiments were carried out with WT ZmC$_4$-NADP-ME and F140I in 10 mM Tris-HCl (pH 8.0) and 5 mM MgCl$_2$ using a Beckman Optima XL-A analytical ultracentrifuge. Amicon Ultra 0.5-ml centrifugal filters with a nominal molecular weight limit of 50 kDa (Merck) were used for buffer exchange procedures. Before centrifugation, protein concentration was adjusted to 0.6 mg ml^{-1}. Samples (230 µg) and buffer solutions (400 µl) were loaded separately into aluminium double-sector centrepieces and built up in a Beckman An-50 Ti rotor. Experiments were performed at 20 °C and 35,000 r.p.m. Protein samples were monitored at 280 nm in continuous mode with a radial resolution of 0.003 cm. In time intervals of about 2 min, scans of the radial concentration profile were collected until the protein was fully sedimented. Data were analysed using the ls-g*(s) model in the software package SEDFIT (v.15.01b)[54]. For data analysis, a resolution of 0.1 S with a confidence level (F-ratio) of 0.683 was chosen for the appropriate s-value range within 0–100 S in the case of the WT, and 0–500 S in case of the F140I variant. The density and viscosity of the solvent were calculated with the software Sednterp (v.20130828 BETA, supported by the Biomolecular Interaction Technologies Center at the University of New Hampshire) from tabulated values: $\rho = 0.99885\ \mathrm{g\ cm}^{-3}$ and $\eta = 0.01006\ \mathrm{g\ cm}^{-1}\ \mathrm{s}^{-1}$. The protein partial specific volume applied for WT and mutants was 0.7387 cm^3 g^{-1}. Graphic output was generated by Gussi (v.1.2.1)[55], and the final sedimentation coefficient distribution was normalized based on maximum peak height.

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The data supporting the findings of this manuscript are available from the corresponding author upon reasonable request. The protein crystallographic structures were deposited in the Protein Data Bank (wwpdb) under the accession codes 5OU5 (ZmC_4-NADP-ME) and 6C7N (SbC_4-NADP-ME).

References

1. 1.

   Amthor, J. S. From sunlight to phytomass: on the potential efficiency of converting solar radiation to phyto-energy. *New Phytol.* **188**, 939–959 (2010).

   - CAS
   - Article
   - Google Scholar
2. 2.

   Osborne, C. P. & Sack, L. Evolution of C4 plants: a new hypothesis for an interaction of CO_2 and water relations mediated by plant hydraulics. *Philos. Trans. R. Soc. Lond. B* **367**, 583–600 (2012).

   - CAS
   - Article
   - Google Scholar
3. 3.

   Sage, R. F., Sage, T. L. & Kocacinar, F. Photorespiration and the evolution of C-4 photosynthesis. *Annu. Rev. Plant Biol.* **63**, 19–47 (2012).

   - CAS
   - Article
   - Google Scholar
4. 4.

   Sage, R. F., Christin, P. A. & Edwards, E. J. The C(4) plant lineages of planet Earth. *J. Exp. Bot.* **62**, 3155–3169 (2011).

   - CAS
   - Article
   - Google Scholar
5. 5.

   Emms, D. M., Covshoff, S., Hibberd, J. M. & Kelly, S. Independent and parallel evolution of new genes by gene duplication in two origins of C4 photosynthesis provides new insight into the mechanism of phloem loading in C4 species. *Mol. Biol. Evol.* **33**, 1796–1806 (2016).

   - CAS
   - Article
   - Google Scholar
6. 6.

Maier, A., Zell, M. B. & Maurino, V. G. Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C(4) and C(3) photosynthesis. *J. Exp. Bot.* **62**, 3061–3069 (2011).

- CAS
- Article
- Google Scholar

7. 7.

Maurino, V. G., Saigo, M., Andreo, C. S. & Drincovich, M. F. Non-photosynthetic ‘malic enzyme’ from maize: a constituvely expressed enzyme that responds to plant defence inducers. *Plant Mol. Biol.* **45**, 409–420 (2001).

- CAS
- Article
- Google Scholar

8. 8.

Saigo, M. et al. Maize recombinant non-C4 NADP-malic enzyme: a novel dimeric malic enzyme with high specific activity. *Plant Mol. Biol.* **55**, 97–107 (2004).

- CAS
- Article
- Google Scholar

9. 9.

Tausta, S. L., Coyle, H. M., Rothermel, B., Stiefel, V. & Nelson, T. Maize C4 and non-C4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. *Plant Mol. Biol.* **50**, 635–652 (2002).

- CAS
- Article
- Google Scholar

10. 10.

Christin, P. A., Samaritani, E., Petitpierre, B., Salamin, N. & Besnard, G. Evolutionary insights on C4 photosynthetic subtypes in grasses from genomics and phylogenetics. *Genome Biol. Evol.* **1**, 221–230 (2009).

- Article
- Google Scholar

11. 11.

Saigo, M., Alvarez, C. E., Andreo, C. S. & Drincovich, M. F. Plastidial NADP-malic enzymes from grasses: unraveling the way to the C4 specific isoforms. *Plant Physiol. Biochem.* **63**, 39–48 (2013).

- CAS
- Article
- Google Scholar

12. 12.

Ashton, A. R. NADP-malic enzyme from the C4 plant *Flaveria bidentis*: nucleotide substrate specificity. *Arch. Biochem. Biophys.* **345**, 251–258 (1997).

- CAS
- Article
- Google Scholar

13. 13.

Iglesias, A. A. & Andreo, C. S. Kinetic and structural properties of NADP-malic enzyme from sugarcane leaves. *Plant Physiol.* **92**, 66–72 (1990).

- CAS
- Article
- Google Scholar

14. 14.

Detarsio, E., Wheeler, M. C., Campos Bermudez, V. A., Andreo, C. S. & Drincovich, M. F. Maize C4 NADP-malic enzyme. Expression in *Escherichia coli* and characterization of site-directed mutants at the putative nucleoside-binding sites. *J. Biol. Chem.* **278**, 13757–13764 (2003).

- CAS
- Article
- Google Scholar

15. 15.

Detarsio, E., Alvarez, C. E., Saigo, M., Andreo, C. S. & Drincovich, M. F. Identification of domains involved in tetramerization and malate inhibition of maize C4-NADP-malic enzyme. *J. Biol. Chem.* **282**, 6053–6060 (2007).

- CAS
- Article
- Google Scholar

16. 16.

Werdan, K., Heldt, H. W. & Milovancev, M. The role of pH in the regulation of carbon fixation in the chloroplast stroma. Studies on CO_2 fixation in the light and dark. *Biochim. Biophys. Acta* **396**, 276–292 (1975).

- CAS
- Article
- Google Scholar

17. 17.

Fahnenstich, H. et al. Alteration of organic acid metabolism in *Arabidopsis* overexpressing the maize C(4)NADP-malic enzyme causes accelerated senescence during extended darkness. *Plant Physiol.* **145**, 640–652 (2007).

- CAS
- Article
- Google Scholar

18. 18.

    Zell, M. B. et al. Analysis of *Arabidopsis* with highly reduced levels of malate and fumarate sheds light on the role of these organic acids as storage carbon molecules. *Plant Physiol.* **152**, 1251–1262 (2010).

    - CAS
    - Article
    - Google Scholar

19. 19.

    Matsumura, H. et al. Crystal structures of C4 form maize and quaternary complex of *E. coli* phosphoenolpyruvate carboxylases. *Structure* **10**, 1721–1730 (2002).

    - CAS
    - Article
    - Google Scholar

20. 20.

    Blasing, O. E., Westhoff, P. & Svensson, P. Evolution of C4 phosphoenolpyruvate carboxylase in *Flaveria*, a conserved serine residue in the carboxyl-terminal part of the enzyme is a major determinant for C4-specific characteristics. *J. Biol. Chem.* **275**, 27917–27923 (2000).

    - CAS
    - PubMed
    - Google Scholar

21. 21.

    Engelmann, S., Blasing, O. E., Westhoff, P. & Svensson, P. Serine 774 and amino acids 296 to 437 comprise the major C4 determinants of the C4 phosphoenolpyruvate carboxylase of *Flaveria trinervia*. *FEBS Lett.* **524**, 11–14 (2002).

    - CAS
    - Article
    - Google Scholar

22. 22.

    Jacobs, B., Engelmann, S., Westhoff, P. & Gowik, U. Evolution of C(4) phosphoenolpyruvate carboxylase in *Flaveria*: determinants for high tolerance towards the inhibitor l-malate. *Plant Cell Environ.* **31**, 793–803 (2008).

    - CAS
    - Article
    - Google Scholar

23. 23.

    Paulus, J. K., Schlieper, D. & Groth, G. Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nat. Commun.* **4**, 1518 (2013).

    - Article

- Google Scholar

24. 24.

    Xu, Y., Bhargava, G., Wu, H., Loeber, G. & Tong, L. Crystal structure of human mitochondrial NAD(P)$^+$-dependent malic enzyme: a new class of oxidative decarboxylases. *Structure* **7**, 877–889 (1999).

    - CAS
    - Article
    - Google Scholar

25. 25.

    Yang, Z., Floyd, D. L., Loeber, G. & Tong, L. Structure of a closed form of human malic enzyme and implications for catalytic mechanism. *Nat. Struct. Biol.* **7**, 251–257 (2000).

    - CAS
    - Article
    - Google Scholar

26. 26.

    Yang, Z. et al. Potent and competitive inhibition of malic enzymes by lanthanide ions. *Biochem. Biophys. Res. Commun.* **274**, 440–444 (2000).

    - CAS
    - Article
    - Google Scholar

27. 27.

    Yang, Z., Lanks, C. W. & Tong, L. Molecular mechanism for the regulation of human mitochondrial NAD(P)$^+$- dependent malic enzyme by ATP and fumarate. *Structure* **10**, 951–960 (2002).

    - CAS
    - Article
    - Google Scholar

28. 28.

    Tao, X., Yang, Z. & Tong, L. Crystal structures of substrate complexes of malic enzyme and insights into the catalytic mechanism. *Structure* **11**, 1141–1150 (2003).

    - CAS
    - Article
    - Google Scholar

29. 29.

    Yang, Z. et al. Structural studies of the pigeon cytosolic NADP($^+$)-dependent malic enzyme. *Protein Sci.* **11**, 332–341 (2002).

    - CAS
    - Article

- Google Scholar

30. 30.

    Alvarez, C. E. et al. The crystal structure of the malic enzyme from *Candidatus Phytoplasma* reveals the minimal structural determinants for a malic enzyme. *Acta Crystallogr. D* **74**, 332–340 (2018).

- 31.

Chang, G. G. & Tong, L. Structure and function of malic enzymes, a new class of oxidative decarboxylases. *Biochemistry* **42**, 12721–12733 (2003).

- CAS
- Article
- Google Scholar

- 32.

Roalson, E. H. in *C4 Photosynthesis and Related CO $_2$ Concentrating Mechanisms* (eds Raghavendra, A. S. & Sage, R. F.) 319–338 (Kluwer, 2011).

- 33.

Hsieh, J. Y., Chen, S. H. & Hung, H. C. Functional roles of the tetramer organization of malic enzyme. *J. Biol. Chem.* **284**, 18096–18105 (2009).

- CAS
- Article
- Google Scholar

- 34.

Izui, K., Matsumura, H., Furumoto, T. & Kai, Y. Phosphoenolpyruvate carboxylase: a new era of structural biology. *Annu. Rev. Plant Biol.* **55**, 69–84 (2004).

- CAS
- Article
- Google Scholar

- 35.

Maurino, V. G. & Weber, A. P. Engineering photosynthesis in plants and synthetic microorganisms. *J. Exp. Bot.* **64**, 743–751 (2013).

- CAS
- Article
- Google Scholar

- 36.

Covshoff, S. & Hibberd, J. M. Integrating C4 photosynthesis into C3 crops to increase yield potential. *Curr. Opin. Biotechnol.* **23**, 209–214 (2012).

- CAS
- Article
- Google Scholar

- 37.

Hibberd, J. M., Sheehy, J. E. & Langdale, J. A. Using C-4 photosynthesis to increase the yield of rice –rationale and feasibility. *Curr. Opin. Plant Biol.* **11**, 228–231 (2008).

- CAS
- Article
- Google Scholar

- 38.

Kajala, K. et al. Strategies for engineering a two-celled C(4) photosynthetic pathway into rice. *J. Exp. Bot.* **62**, 3001–3010 (2011).

- CAS
- Article
- Google Scholar

- 39.

Goodstein, D. M. et al. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186 (2012).

- CAS
- Article
- Google Scholar

- 40.

Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc. Natl Acad. Sci. USA* **102**, 10557–10562 (2005).

- Article
- Google Scholar

- 41.

Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

- CAS
- Article
- Google Scholar

- 42.

Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).

  - CAS
  - Article
  - Google Scholar

- 43.

Battye, T. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D* **67**, 271–281 (2011).

  - CAS
  - Article
  - Google Scholar

- 44.

Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).

  - CAS
  - Article
  - Google Scholar

- 45.

Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).

  - CAS
  - Article
  - Google Scholar

- 46.

Bricogne, G. et al. *BUSTER v.2.9* (Global Phasing Ltd, 2009).

- 47.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).

  - CAS
  - Article
  - Google Scholar

- 48.

The PyMOL Molecular Graphics System (DeLano Scientific, 2002).

- 49.

Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).

  - CAS
  - Article
  - Google Scholar

- 50.

van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R. & Groen, F. C. Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal. Biochem.* **191**, 110–118 (1990).

  - Article
  - Google Scholar

- 51.

Provencher, S. W. & Glockner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20**, 33–37 (1981).

  - CAS
  - Article
  - Google Scholar

- 52.

Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260 (2000).

  - CAS
  - Article
  - Google Scholar

- 53.

Whitmore, L. & Wallace, B. A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668–W673 (2004).

  - CAS
  - Article
  - Google Scholar

- 54.

Schuck, P. & Rossmanith, P. Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* **54**, 328–341 (2000).

- CAS
- Article
- Google Scholar

- 55.

Brautigam, C. A. Calculations and publication-quality illustrations for analytical ultracentrifugation data. *Methods Enzymol.* **562**, 109–133 (2015).

Acknowledgements

This work was funded by grants of the European Union (3to4) to V.G.M. and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy (EXC 2048/1, Project ID: 390686111 and EXC 1028, to V.G.M. and M.J.L. We acknowledge the European Synchrotron Radiation Facility (ESRF) for provision of synchrotron radiation facilities and we wish to thank L. Gordon for assistance in using beamline ID23-1. The Center for Structural Biology of Mercosur granted funding to C.E.A. for data collection at Institut Pasteur de Montevideo (IPM). We are grateful to N. Larrieux at the Protein Crystallography Facility IPM for assistance with crystallization and data collection.